May 2008

Viral Marketing and the Diffusion of Trends on Social Networks

Jennifer Wortman
University of Pennsylvania, wortmanj@seas.upenn.edu

University of Pennsylvania Department of Computer and Information Science Technical Report No. MS-CIS-08-19.

Viral Marketing and the Diffusion of Trends on Social Networks

Abstract

We survey the recent literature on theoretical models of diffusion in social networks and the application of these models to viral marketing. To put this work in context, we begin with a review of the most common models that have been examined in the economics and sociology literature, including local interaction games, threshold models, and cascade models, in addition to a family of models based on Markov random fields. We then discuss a series of recent algorithmic and analytical results that have emerged from the computer science community. The first set of results addresses the problem of influence maximization, in which the goal is to determine the optimal group of individuals in a social network to target with an advertising campaign in order to cause a new product or technology to spread throughout the network. We then discuss an analysis of the properties of graphs that allow or prohibit the widespread propagation of trends.

Comments

University of Pennsylvania Department of Computer and Information Science Technical Report No. MS-CIS-08-19.

Viral Marketing and the Diffusion of Trends on Social Networks

Jennifer Wortman

Technical Report MS-CIS-08-19
Department of Computer and Information Science
University of Pennsylvania

May 15, 2008

Abstract

We survey the recent literature on theoretical models of diffusion in social networks and the application of these models to viral marketing. To put this work in context, we begin with a review of the most common models that have been examined in the economics and sociology literature, including local interaction games, threshold models, and cascade models, in addition to a family of models based on Markov random fields. We then discuss a series of recent algorithmic and analytical results that have emerged from the computer science community. The first set of results addresses the problem of influence maximization, in which the goal is to determine the optimal group of individuals in a social network to target with an advertising campaign in order to cause a new product or technology to spread throughout the network. We then discuss an analysis of the properties of graphs that allow or prohibit the widespread propagation of trends.

1 Introduction

For decades, social scientists have been interested in the fundamental question of how new trends, behaviors, and innovations spread through social networks. Models of network diffusion have been used to study phenomena as widespread as product recommendation systems and viral marketing [41, 20, 53, 43, 44], information propagation on blogs [45, 46, 31, 1], the transmission of infectious diseases or computer viruses [18, 19, 3], the prevention of water contamination [45], herding behavior in financial markets [58, 23], diffusion of government policies such as anti-smoking laws or state lottery adoption [56, 4], the spread of new agricultural or medical practices [54, 13], Hollywood trends [17, 16], and even the contagion properties of obesity [11].

With the recent growth of the Internet, it is now possible to observe diffusion on huge social networks gathered from online recommendation systems, social networking websites, and instant messaging systems. These networks are significantly larger than any that have been studied before, in some cases containing hundreds of millions of nodes and more than a billion edges [42]. As such, computer scientists have recently become interested in developing techniques for analyzing the diffusion properties of these large-scale networks, where efficiency is a real concern. Such analyses are immediately applicable in the area of viral marketing.

Viral marketing hinges on the premise that a consumer's purchasing decisions are heavily influenced by recommendations and referrals from his family, friends, and colleagues. Sometimes the desired network effects are explicit in the marketing efforts. For example, when the Hotmail web-based email service started in 1996, each message sent by a Hotmail user included a promotional message with the URL of the service. As a result, Hotmail gained its first twelve million subscribers in just eighteen months, on an advertising budget of only $50,000 [33]. In other cases, products spread from friend to friend due to compatibility issues. Cell phone service providers such as AT&T and Sprint offer "mobile-to-mobile" deals in which customers may talk to other customers of the same provider for free or at heavily discounted rates. If someone's family and closest friends are all AT&T customers, there is an added incentive for him to sign up for AT&T. Similarly, if someone's colleagues primarily edit documents in Latex, it may be easier for him to switch technologies than to insist on using Word.

In this paper, we provide a survey of the recent literature on theoretical models of diffusion in social networks, and discuss the application of these models to viral marketing and, more generally, to the study of how new products or technologies spread through a population. We begin with an overview of models that have been examined in the economics and sociology literature, including local interaction games [7, 24, 28], threshold models [30, 55], and cascade models [26, 27, 36], as well as a family of models based on Markov random fields [39, 6]. With this framework in place, we move on to algorithmic and analytical results from the computer science community.

The first algorithmic problem we discuss is the problem of influence maximization in a viral marketing setting. Here we are given access to a social network and the goal is to find the optimal group of individuals to target for marketing efforts in order to trigger the widespread adoption of a new product or innovation. Because of the potential complexity of the underlying structure of the social network, obvious heuristic approaches such as always choosing the individuals with the highest degree in the network are generally not optimal. In fact, as we will see in Section 3.1, Kempe et al. [36] show that in a variety of natural settings, solving the influence maximization problem exactly is NP-hard. However, they provide a very simple greedy algorithm that can be used to obtain approximate results in many of these settings. Domingos and Richardson [21, 53] take a different approach, modeling the social network as a Markov random field, and provide a variety of heuristics that can be used to select a group of individuals to target for advertising. This approach is discussed in Section 3.2.

In Section 4, we move on to the related but distinct problem of analyzing the properties of graphs that allow or prohibit the widespread contagion of products. We begin with an overview of the seminal work of Morris [48], who provided an extensive analysis of the graph properties related to contagion when modeling the spread of a new product using a local interaction game. We then discuss a recent extension of this work by Immorlica et al. [32], who consider a scenario in which individuals are not restricted to either using only the old "incumbent" technology or the new technology that is spreading around the network, but may in fact decide to use a combination of both at some added cost. For example, if half of someone's colleagues edit documents in Latex, and the other half use Word, it might be advantageous for the person to use both

technologies, despite the potential hassle of switching back and forth.

We conclude with some remarks about potential directions for further research in Section 5.

2 Social Networks and Models of Diffusion

Empirical studies of diffusion on social networks began to appear in the middle of the 20th century [54, 13], but formal mathematical models of diffusion were not introduced until decades later in work by Granovetter [30] and Schelling [55]. In this section, we survey a variety of the mathematical models for network diffusion that have emerged from the economics and sociology communities in recent years. We begin with some simple definitions that are common to all of the models we will examine.

Formally, a *social network* is represented as a graph $\mathcal{G} = (V, E)$, which can be either directed or undirected depending on the particular model and application. Each vertex $v \in V$ represents an individual in the network. For convenience, throughout this paper we refer to the nodes in the graph and the individuals they represent interchangeably. In a directed graph, an edge $(u, v) \in E$ signifies that u has direct influence on the decisions made by v. Similarly, in an undirected graph, the edge $(u, v) = (v, u)$ signifies mutual influence between u and v. Finally, we denote by $\mathcal{N}(v)$ the neighbors of v. This set contains all individuals who have direct influence on v. More precisely, $\mathcal{N}(v) = \{u \in V : (u, v) \in E\}$.

We now introduce several models of diffusion, each of which makes different assumptions on the ways in which a node is influenced by his neighbors.

2.1 Local Interaction Games

Consider two drivers approaching each other along a narrow dirt road. In order to avoid collision, the drivers must swerve. If both swerve to the right, they will successfully avoid each other. Similarly, if both swerve to the left, they will be fine. However, if one driver swerves left while the other swerves right, their cars will collide.

This scenario is a (somewhat extreme) example of a simple coordination game. A coordination game is a two-player symmetric game in which it is in the best interest of both players to choose identical actions. If one player chooses action A, then A is the best response of the other; if one player chooses action B, then B is the best response. A sample payoff matrix for a coordination game is given in Table 1. Each row of the matrix corresponds to an action of Player 1, while each column corresponds to an action of Player 2. The two numbers in each square of the matrix give the payoffs of Player 1 and Player 2 respectively. (In this case, the payoffs to both players are identical.) Here $q \in (0, 1)$ is a parameter determining the relative goodness of action A compared to action B.

	A	B
A	$1 - q$, $1 - q$	0, 0
B	0, 0	q, q

Table 1: A simple two-player coordination game.

Perhaps a more realistic example of a coordination game is that of choosing between competing technologies. For example, consider two friends who would like to chat online. In order to chat, both need to install software. If both install AOL Instant Messenger, they can use this software to chat with each other. Similarly, if both install Yahoo! Messenger, they can use this software to chat. However, if one installs AOL Instant Messenger while the other installs Yahoo! Messenger, they cannot chat with each other due to the incompatibility of the software.

This scenario gets more interesting when we consider that these two friends are probably part of a much larger social network. Suppose each individual in the network would like to chat with all of his neighbors. If his neighbors unanimously decide to install Yahoo! Messenger, then it is clearly a best response for the

individual to install Yahoo! Messenger too. When different subsets of his neighbors choose different chat clients, the best response is less clear.

Local interaction games were developed as an extension of two-player coordination games to the setting in which there is a large or even infinite population of players interacting along the edges of a social network. Variations of local interaction games were presented and studied in the early nineties by Blume [7] and Ellison [24], and later by Goyal [28], and they can be viewed as a precursor to the more general, widely studied class of graphical games [35, 34]. In this paper, we focus primarily on the version of local interaction games formalized by Morris [48].

For our purposes, a local interaction game is a pair $(\mathcal{G}, q)$ where $\mathcal{G} = (V, E)$ is a connected, undirected graph representing a social network. The vertex set V and edge set E may be finite or infinite, but it is assumed that no node has more than a finite number M of neighbors. Here, as above, q is parameter specifying the relative goodness of the two actions A and B. In particular, if a player chooses action B, then he receives a payoff of q for *each* of his neighbors choosing B. If a player chooses A, he receives a payoff of $1-q$ for each neighbor choosing A. Thus we can think of each player as interacting with each of his neighbors via the matrix game shown in Table 1. The payoff to the player is the sum of the payoffs of each of these games. The only restriction is that each player must choose the *same* action for each game.

Many Nash equilibria exist for this game. For example, the situation in which all players choose action A is an equilibrium, as is the situation in which all players choose action B. Examining a single equilibrium does not give much insight about how trends spread. Instead, in order to model the diffusion of technology over time, it is more useful to consider the evolution of the players' actions over time if players are assumed to play best responses at each time step.

Specifically, assume that the (possibly infinite) set of nodes V start out in a particular configuration at time 1, with a set of players $\mathcal{A}_1$ choosing A and the remaining players $V \setminus \mathcal{A}_1$ choosing B. At each subsequent time t, each player plays a best response to the actions chosen by his neighbors at time $t-1$. (To be concrete, we assume that ties are broken in favor of action A.) We generally think of choosing action B as choosing the old, "incumbent" technology, or as choosing not to adopt any technology at all. Conversely, we think of choosing action A as choosing to adopt the "new" technology. Thus when the graph is finite, we generally care only about cases in which the set $\mathcal{A}_1$ is small compared to V. In the same spirit, when the graph is infinite, we restrict our attention to the case in which $\mathcal{A}_1$ is finite. We say that a node v is *active* if v is currently choosing action A, and *inactive* otherwise, and let $\mathcal{A}_t$ refer to the set of active nodes at time t. Examining the evolution of the active set over time yields a dynamic view of how the new product or technology may spread across the network.

2.2 Threshold Models

Local interaction games can be viewed as a simple subclass of the more general class of threshold models. Recall that in a local interaction game, a node v chooses his action at time t as a best response to his neighbors actions at time $t-1$. If v chooses action A, his payoff is $1-q$ times the number of his neighbors who choose A; if he chooses action B, his payoff is q times the number of his neighbors who choose B. Thus A is a best response for v if and only if a fraction q or more of his neighbors choose action A. Formally, node v chooses action A at time t if and only if

$$\frac{1}{|\mathcal{N}(v)|} \sum_{u \in \mathcal{N}(v)} X_{u,t-1} \geq q \ ,$$

where $X_{u,t-1}$ is 1 if u chose action A at time $t-1$ and 0 otherwise. Here the parameter q can be viewed as a fixed cutoff or *threshold*.

Threshold models date back at least as far as Granovetter [30] and Schelling [55]. One simple example is the *linear threshold model*. In this model, each node $v \in V$ has a nonnegative weight $\omega_{v,u}$ for every $u \in \mathcal{N}(v)$, where $\sum_{u \in \mathcal{N}(v)} \omega_{v,u} \leq 1$, and a personal threshold value θ_v. This threshold can be hard-wired to a particular value as in the case of local interaction games, or chosen at random at the start of the process. Given these thresholds and an initial set $\mathcal{A}_1$ of active nodes, the process unfolds deterministically in a sequence of steps.

At the tth time step, every node that was active at time $t-1$ automatically remains active. Each node v that was inactive at time $t-1$ becomes active at time t if and only if

$$\sum_{u \in \mathcal{N}(v)} \omega_{v,u} X_{u,t-1} \geq \theta_v \ ,$$

where $X_{u,t-1}$ is 1 if u was active at time $t-1$ and 0 otherwise. Intuitively then, the weight $\omega_{v,u}$ represents the extent to which v is influenced by u, and the threshold θ_v represents the personal tendency of v to adopt a new technology when his neighbors do.

It is important to note that while the process is deterministic *given the set of thresholds*, the ability to randomize the choice of thresholds allows a natural injection of a source of randomness into the process. For example, Kempe et al. [36, 37] assume that the threshold is chosen independently for each player, uniformly at random in $(0, 1)$, in order to model the lack of prior knowledge about the true thresholds of each individual. This ability to inject randomness is critical for some of the results that follow, especially the equivalence results in Section 3.1.

The linear threshold model can be generalized further in a very natural way by replacing the term $\sum_{u \in \mathcal{N}(v)} \omega_{v,u} X_{u,t-1}$ with an arbitrary function of the set of active neighbors of v [36]. More specifically, let f_v be any monotone increasing function with $f_v(\emptyset) = 0$ that maps (active) subsets of $\mathcal{N}(v)$ to the range $[0, 1]$. Similar to the linear threshold model, in the general model at each time t, each inactive node v becomes active if and only if $f_v(S) \geq \theta_v$, where $S = \mathcal{A}_{t-1} \cap \mathcal{N}(v)$ is the subset of $\mathcal{N}(v)$ that were active at time $t-1$. As before, thresholds may be fixed or chosen randomly at the start of the process; throughout this paper, we generally assume that the thresholds are chosen independently and uniformly at random in $(0, 1)$ for each $v \in V$.

Note that in this model, once a node has switched to action A, he never switches back to action B. This property is known as *progressiveness*. Most of the dynamic processes we examine in this paper are progressive, either explicitly as in this model, or implicitly as a result of the model dynamics.

2.3 Cascade Models

Inspired by research on interacting particle systems [47], *cascade models* of diffusion were first studied in the context of marketing by Goldenberg et al. [26, 27]. In the cascade model, each individual has a single, probabilistic chance to activate each inactive node for which he is a neighbor after becoming active himself. A very simple example is the *independent cascade model*, in which the probability that an individual is activated by a newly active neighbor is independent of the set of neighbors who have attempted to activate him in the past. Once again, starting with an initial active set $\mathcal{A}_1$, the process unfolds in a series of time steps. At each time t, any node v who has just become active may attempt to activate each inactive node u for which $v \in \mathcal{N}(u)$. With probability $p_{u,v}$, u becomes active at the next time step. Whether or not u becomes active, v and u have no further contact throughout the remainder of the process.

It is also simple to generalize the independent cascade model. Instead of node v activating u with probability $p_{u,v}$, node v activates u with a probability that depends on the set of nodes who have tried to activate u in the past. Formally, we can define a probability $p_{u,v}(S)$ for each $S \in \mathcal{N}(u) \setminus \{v\}$. When node v first becomes active, he succeeds in activating u with probability $p_{u,v}(S)$, where S is the set of neighbors of u who have previously tried (and failed) to activate u. In order to ensure that the model is always well-defined, we may consider only sets of probabilities $p_{u,v}(S)$ that are *order-independent*, so the probability that an individual u is active after a set U of his neighbors have tried to activate him does not depend on the order in which the neighbors U made their attempts; without this property, it would not be clear what should happen in the case that multiple neighbors of a node become active on the same round. As we will see in Section 3.1.3, this general class of cascade models is equivalent to the general class of threshold models in a very strong sense.

By definition, the independent cascade model and the generalized cascade model are also progressive.

2.4 Markov Random Fields

Local interaction games, more general threshold models, and cascade models all assume a particular structure on the way in which each node chooses whether or not to adopt action A given the choices of his neighbors. Furthermore, these models are all *dynamic* in that they explicitly model the evolution of the actions over time. An alternative to modeling the progression of actions in this way is to model only the final state of the network at convergence as one large global set of interdependent random variables. This can be accomplished by modeling the actions of each node in the social network as a Markov random field.

A *Markov network*, or *Markov random field*, is an undirected graphical model representing the joint distribution over a set of random variables. Each node of the Markov network represents a single variable. Each edge represents a dependence between variables.

Markov networks satisfy a variety of properties that make them useful for reasoning about large, structured domains. For example, they provide a simple method for determining whether two variables, X and Y, are conditionally independent given a set of variables U. In particular, it is sufficient to check the set of paths between X and Y. If each path passes through at least one node in U, then X and Y are conditionally independent; otherwise, they are not. Inference algorithms have been designed to take advantage of the independence properties and graphical structure of Markov networks. The belief propagation algorithm allows efficient inference when the underlying network is a tree, and multiple extensions exist for efficient approximate inference in general graphs. For a more detailed overview, see Kindermann and Snell [39] or Chapter 8 of Bishop [6].

Consider a social network $\mathcal{G} = (V, E)$. For each $v \in V$, we define a Boolean variable X_v that corresponds to whether or not v adopts the product; it takes on the value 1 if v is active and 0 otherwise. These new variables naturally form a Markov network with the same edge structure as the original social network. This representation is powerful in that it allows the probability that a node v adopts action A to depend on whether or not his neighbors adopt A in an arbitrary way.

In order to use this representation to reason about marketing decisions, additional nodes can be added to the Markov network representing, for example, the extent to which each player is targeted for marketing. These ideas are discussed in more detail in Section 3.2.

3 Maximizing Influence

In this section, we discuss influence maximization. Influence maximization is the problem of choosing the "best" set of individuals in a network to market to in order to trigger the widespread adoption of a product. This problem was first introduced by Domingos and Richardson [21], who noted that while data mining techniques had been used in various direct marketing applications with great success, a wealth of network information was ignored in these traditional applications. Instead of reasoning about the buying decisions of each consumer independently when determining who to target for marketing, they suggested incorporating a customer's network effect into their value. Thus instead of deciding whether or not to market to a consumer based solely on the expected profit that would be gained from the particular consumer making a purchase, marketing companies should instead take into account the effect that this consumer would have if she told her friends, family, and colleagues about the product and they, in turn, told their own friends.

The influence maximization problem can be formalized in a number of different ways. Domingos and Richardson [21, 53] model the problem as a Markov random field, and discuss heuristic techniques aimed at finding a marketing strategy that (approximately) maximizes the the global *expected lift in profit*. Intuitively, the expected lift in profit, which was introduced by Chickering and Heckerman [10] in the context of direct marketing, is the difference between the expected profit obtained by employing a marketing strategy and the expected profit obtained using no marketing at all.

In more recent work, Kempe et al. [36, 37] assume a fixed marketing budget sufficient to target k individuals and study the problem of finding the optimal k individuals in the network to target. They show that in a variety of very simple models, including the linear threshold model and independent cascade model, this problem is NP-hard, but provide a straight-forward greedy algorithm, analogous to the standard greedy

set cover approximation algorithm [14], that is guaranteed to efficiently produce a $(1 - 1/\mathrm{e} - \epsilon)$-optimal set for any $\epsilon > 0$.

We discuss these models in turn, and then provide a brief comparison of the applicability of each.

3.1 k-Best Influence Maximization

We begin by discussing the variant of the influence maximization problem posed by Kempe et al. [36, 37]. Consider a social network $\mathcal{G}$ in which each node chooses his action at each time step according to some function of the actions chosen in his local neighborhood at the previous time step. For example, nodes might choose their actions according to the general threshold or cascade model described in Section 2. We assume that $\mathcal{G}$ is directed, but all of the results in this section apply to undirected graphs as well.

Formally, Kempe et al. define the *influence* of a set of nodes $\mathcal{A}$, denoted $\sigma(\mathcal{A})$, as the expected number of nodes active at the end of a process given that the initial set of active nodes $\mathcal{A}_1$ is set to $\mathcal{A}$. The *k-best influence maximization problem*[1] is then the problem of finding the set $\mathcal{A}$ of a fixed size k that maximizes $\sigma(\mathcal{A})$. Here it is assumed that the network itself and the parameters of each player (e.g., the threshold functions in the threshold model, or the activation probabilities in the cascade model) are known and given to the algorithm as input.

In general, the k-best influence maximization problem is NP-hard, even in very restricted settings. However, Kempe et al. show that it is possible to approximate the solution in a variety of special cases using a simple greedy algorithm. In the next section, we illustrate the hardness of the problem for the independent cascade model. We then introduce a more general subclass of cascade models, decreasing cascade models, and show that k-best influence maximization can be approximated for this class. The proof hinges on the fact that the influence function is *submodular* for this class and uses the equivalence between the general threshold and general cascade models, which is discussed in more detail below.

3.1.1 Hardness

Recall the independent cascade model introduced in Section 2. In this model, when a node v becomes active, he activates each of the inactive nodes u who have him as a neighbor with probability $p_{u,v}$, independent of the other nodes who have tried to activate u in the past. The following theorem, proved by Kempe et al. [36], shows that the k-best influence maximization problem is NP-hard, even when restricted to this very simple model.

Theorem 1 *The k-best influence maximization problem is NP-hard, even when restricted to the independent cascade model.*

Proof: The proof is via reduction from the NP-complete set cover problem. In this problem, we are given a collection subsets $S_1, S_2, \cdots, S_m$ of a base set $U = \{u_1, u_2, \cdots, u_n\}$ and a fixed integer k, and wish to determine whether or not there exist k of these subsets whose union is the full set U.

Starting with an arbitrary instance of the set cover problem, it is possible to construct a corresponding instance of k-best influence maximization as follows. First, create a node corresponding to each set S_i for $i \in \{1, \cdots, m\}$, and a node corresponding to each element $u_k \in U$. For each $i \in \{1, \cdots, m\}$ and $j \in \{1, \cdots, n\}$, create a directed edge from the node corresponding to S_i to the node corresponding to u_j with activation probability 1 if and only if $u_j \in S_i$. An example of such a graph is shown in Figure 1.

It remains to show that given access to an oracle who could solve the influence maximization problem on this new graph, it would be possible to efficiently compute the answer to the set cover problem. Since the activation probabilities on each edge are 1, the diffusion process deterministic, and it is possible to exactly compute $\sigma(A)$ for any set A by simulating the diffusion process a single time. Thus, if we knew the set A that maximizes influence with parameter k, we could easily determine whether $\sigma(A) \geq n + k$.

[1] We introduce the terminology "k-best" to distinguish this version of the influence maximization problem from the version discussed in Section 3.2.

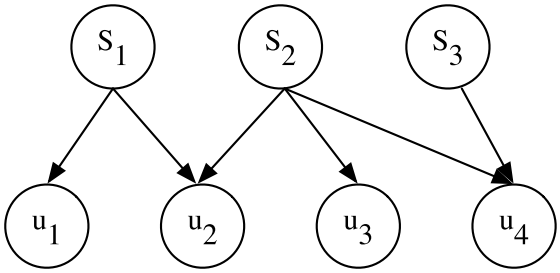

Figure 1: Illustration of the reduction from set cover to the problem of k-best influence maximization under the independent cascade model. Here $S_1 = \{u_1, u_2\}$, $S_2 = \{u_2, u_3, u_4\}$, and $S_3 = \{u_4\}$. Suppose $k = 2$. If each edge has weight 1, the k-element set A of nodes that maximizes $\sigma(A)$ is $\{S_1, S_2\}$, with $\sigma(\{S_1, S_2\}) = 6$. The sets S_1 and S_2 also cover $U = \{u_1, \cdots, u_4\}$.

Suppose $\sigma(A) \geq n + k$. Since the nodes corresponding to the sets S_i have no incoming edges, they can only be active if they are members of the original set A. This implies that there must be at most k active nodes corresponding to sets, and all n nodes corresponding to set members u_i must be active. The only way that these n nodes can be active is if each is contained in at least one of the active set nodes, implying that there exists a set cover of size k. By a similar argument, if $\sigma(A) < n + k$, then no set cover exists. ∎

A similar result is easily achievable for the linear threshold model.

3.1.2 A General Approach to Approximation

While the k-best influence maximization problem is NP-hard, it can be efficiently approximated in many special cases using a very simple and intuitive greedy algorithm, defined below. This algorithm builds the initial active set one element at a time, always greedily choosing the element that results in the largest immediate gain in influence.

Algorithm 1 The greedy k-best influence maximization algorithm.

Initialize $S_0 \leftarrow \emptyset$
for $i = 1$ to k **do**
  Set $S_i \leftarrow S_{i-1} \cup \text{argmax}_{u \in V} \sigma(S_{i-1} \cup \{u\})$
end for
Set $\mathcal{A}_1 \leftarrow S_k$

This algorithm is directly analogous to the standard greedy set cover approximation algorithm [14]. In fact, both result in the same sets of active nodes being chosen in the example in Figure 1. The fact that this algorithm often yields a good approximation for the k-best influence maximization problem follows directly from a useful result about submodular functions. Recall that a submodular function is defined as follows.

Definition 1 *Let f be a function that maps subsets of a finite ground set U to nonnegative real numbers. The function f is* submodular *if for any pairs of sets S and T such that $S \subseteq T \subseteq U$, for any element $u \in U$,*

$$f(S \cup \{u\}) - f(S) \geq f(T \cup \{u\}) - f(T) \ .$$

In other words, a function is submodular if it satisfies a certain "diminishing returns" property. Namely it must be the case that the marginal value of adding a new element to a set S decreases as the set S grows. The following theorem about submodular functions is due to Cornuejols et al. [15] and Nemhauser et al. [50].

Theorem 2 *Let f be a monotone increasing, submodular function mapping subsets of a set U to nonnegative real numbers, and let $T \subseteq U$ be a set. Let S be a set of size k, built using the following greedy procedure:*

1. *Set* $S_0 = \emptyset$.
2. *For* $i = 1$ *to* k, *set* $S_i = S_{i-1} \cup \text{argmax}_{t \in T} f(S_{i-1} \cup t)$.
3. *Set* $S = S_k$.

Finally, let S^* *be the* k*-element subset of* T *maximizing* f. *Then*

$$f(S) \geq (1 - 1/e) f(S^*) \geq 0.63 f(S^*),$$

where e *is the base of the natural logarithm. Furthermore, for any* $\gamma > 0$, *if Step 2 is replaced with a step that sets* $S_i = S_{i-1} \cup u$ *for some* u *such that*

$$f(S_{i-1} \cup u) \geq (1 - \gamma) \max_{t \in T} f(S_{i-1} \cup t),$$

then

$$f(S) \geq (1 - 1/e - \epsilon) f(S^*),$$

where ϵ *is a polynomial function of* γ.

This result yields a straight-forward, widely applicable technique for producing approximation results. Namely, if it is possible to show that for a particular model, the influence function $\sigma(\cdot)$ is monotonic and submodular, then it follows immediately that k-best influence maximization can be approximated within a factor of $(1 - 1/\text{e})$ using the simple greedy algorithm defined above. Additionally, even if $\sigma(\cdot)$ cannot be computed exactly but can be efficiently approximated, then the second half of the lemma shows that it is still possible to use the greedy algorithm to efficiently approximate k-best influence maximization.

Note that submodularity of the influence function is a somewhat natural property to expect in many models. It basically says that the marginal gain in profit of adding another individual to the marketing set decreases as the set grows.

3.1.3 Approximability of Influence Maximization for The Decreasing Cascade Model

We have established that the k-best influence maximization problem is NP-hard, even when restricted to fairly simple models like the the independent cascade model. We now illustrate how the approximation technique described in the previous section can be used to show that in many such special cases, k-best influence maximization can be efficiently approximated. Rather than illustrating this result for the independent cascade model, we show that it is true for the broader, more interesting decreasing cascade model.

The *decreasing cascade model* is a natural subclass of the general cascade model in which it is required that the function $p_{u,v}$ be non-increasing. In other words, it must be the case that $p_{u,v}(S) \geq p_{u,v}(T)$ whenever $S \subseteq T$. At a high level, this restriction can be viewed as modeling the phenomenon that individuals become saturated with marketing over time, so additional referrals from friends yield diminishing returns. We assume these probabilities are *order-independent* in the sense that the probability that an individual u is active after a set U of his neighbors have tried to activate him does not depend on the order in which the neighbors U made their attempts. (If this is not the case, then the probability that a node becomes active following a time step in which more than one of his neighbors are activated is ambiguous.) Note that the independent cascade model is a special case of the decreasing cascade model since $p_{u,v}(S) = p_{u,v}(T)$ for all S and T, so the hardness result of Theorem 1 carries over.

Below we sketch a proof of the following theorem, which states that the influence function is submodular for the decreasing cascade model. Since the influence function can be approximated with arbitrarily high precision via sampling, the results in the previous section imply that for any $\epsilon > 0$, it is possible to find a $(1 - 1/\text{e} - \epsilon)$-approximation to the k-best influence maximization problem for this model in polynomial time. We remark that while this result may not be surprising given the "diminishing returns" flavor of the decreasing cascade model, the proof is quite complex, and is interesting in its own right.

Theorem 3 *For any instance of the decreasing cascade model, the influence function* $\sigma(\cdot)$ *is monotone increasing and submodular.*

It is difficult to reason about the submodularity of the influence function directly due to the inherent randomness of the decreasing cascade model. In particular, the proof of Theorem 3 requires reasoning about the effect of delaying the activation attempt of one node on another for many time steps, which is not straight-forward in the cascade model. To simplify this discussion, Kempe et al. [37] first prove that the general cascade model and the general threshold model are equivalent in a precise technical sense described below. Because of this equivalence, we can freely switch back and forth between the cascade view of the model and the threshold view of the model when reasoning about submodularity.

Recall that in the general threshold model, each node v has a corresponding monotone function f_v mapping subsets of $\mathcal{N}(v)$ to the range $[0, 1]$ and a corresponding threshold θ_v. Node v transitions from inactive to active at time t if and only if $f_v(S) \geq \theta_v$, where S is the subset of $\mathcal{N}(v)$ active at time $t-1$. In the general cascade model, each node v has a corresponding activation function $p_{u,v}(\cdot)$ for each neighbor u, where $p_{u,v}(S)$ is the probability that v activates node u when v first becomes active if $S \subseteq \mathcal{N}(u)$ is the set of nodes that have previously tried and failed to activate u. In this section, we show that if the threshold functions θ_v are chosen independently and uniformly at random, then there is a mapping between threshold model functions f_v and cascade model probability functions $p_{u,v}$ that makes the two models equivalent in the sense that the distributions over final active sets are the same.

The following lemma is mentioned without proof in Kempe et al. [36] and later proved formally by the same authors [37].

Lemma 1 *There exists a bijective mapping between general threshold model threshold functions $f_v(\cdot)$ and general cascade model activation functions $p_{v,u}(\cdot)$ such that if it is assumed that threshold model thresholds are drawn independently and uniformly at random in $(0, 1)$, then for every set of nodes $T \subseteq V$ and every time t, the probability that $\mathcal{A}_t = T$ is the same under the general threshold model with activation functions $f_v(\cdot)$ and the general cascade model with activation functions $p_{v,u}(\cdot)$.*

In particular, given cascade model success probabilities $p_{v,u}(S)$ for all $u \in \mathcal{N}(v)$ and $S \subseteq \mathcal{N}(v)$, we define the corresponding threshold model activation function for v as

$$f_v(S) = 1 - \prod_{i=1}^{r} \left(1 - p_{v,u_i}(\{u_1, \cdots, u_{i-1}\})\right) \ , \tag{1}$$

where $S = \{u_1, \cdots, u_r\}$. It is easy to show via straight-forward algebra that this function is a bijection; given a threshold model activation function $f_v(\cdot)$, we can equivalently define the corresponding probability $p_{v,u}(S)$ for all $u \in \mathcal{N}(v)$ and $S \subseteq \mathcal{N}(v)$ as

$$p_{v,u}(S) = \frac{f_v(S \cup \{u\}) - f_v(S)}{1 - f_v(S)} \ .$$

Kempe et al. show by induction that if Equation 1 is satisfied, then at each time step t, for every pair of subsets $T, T' \subseteq V$, the probability that $\mathcal{A}_t = T$ and $\mathcal{A}_{t+1} = T'$ is the same in each model. Fixing a value of T and summing over all possible T' yields the lemma.

With Lemma 1 in place, we are now ready to sketch the main ideas of the proof of Theorem 3. Intuitively, it is clear that $\sigma(\cdot)$ is monotonic; this can be proved using a simple induction on the time t as the process evolves. Here we focus on the more intricate proof that $\sigma(\cdot)$ is submodular. The proof relies heavily on the equivalence between the threshold model and the cascade model.

Recall that in order to prove that $\sigma(\cdot)$ is submodular, it is necessary to show that for any sets U and U' such that $U \subseteq U' \subseteq V$ and any node $w \in V$,

$$\sigma(U \cup \{w\}) - \sigma(U) \geq \sigma(U' \cup \{w\}) - \sigma(U') \ .$$

This is trivially true if $w \in U$ or $w \in U'$, so assume that this is not the case. Because it is difficult to reason about the quantities $\sigma(U \cup \{w\}) - \sigma(U)$ and $\sigma(U' \cup \{w\}) - \sigma(U')$ directly, Kempe et al. [37] introduce the notion of a "residual process" and show how this process can be used to reason indirectly about the terms

of interest. For any set of nodes $Z \subset V$, the corresponding residual process is a cascade process defined on the set of nodes $V \setminus Z$. In this new process, for any pair of neighbors $v, u \in V \setminus Z$, the probability that node u activates node v when he first becomes active given that the set of nodes S has already tried and failed to activate v is defined as

$$p^Z_{v,u}(S) = p_{v,u}\left(S \cup (Z \cap \mathcal{N}(V))\right) \ ,$$

where $p_{v,u}(\cdot)$ is the activation function in the original decreasing cascade process. In other words, the probability that node u successfully activates node v in the residual process given that the nodes S have already tried and failed to activate v is the same as the probability in the original process that u activates v given that the set of nodes S and any neighbors of v in Z have already tried and failed.

It is useful to introduce some additional notation to reason about the relationships between these processes. For any set $U \subseteq V$, let $\varphi(U)$ be a random variable representing the set of nodes that are active in the decreasing cascade process at convergence if the initial active set $\mathcal{A}_1$ is set to U. Then by definition, $\sigma(U) = \mathrm{E}[|\varphi(U)|]$. Similarly, let $\varphi^Z(U)$ be a random variable representing the set of nodes active at convergence in the residual process corresponding to the set Z if the initial set of active nodes is U, and define $\sigma^Z(U) = \mathrm{E}[|\varphi^Z(U)|]$.

It is additionally useful to introduce a method of reasoning about the difference between the set of nodes that are active at the end of the process on a particular trial and the set of nodes that *would have been* active if a different initial set of active nodes had been chosen. In order to do this, we assume that all of the randomness of the process is fixed at the start of each trial. In the cascade model, this randomness corresponds to the outcomes of the biased coin flips determining whether or not each node activates his neighbor if given the chance. In the threshold model, the randomness corresponds to the threshold value θ_v for each v. For clarity of exposition, we slightly depart from Kempe et al. [37] and introduce a variable R to explicitly represent a particular setting of this randomness. We then abuse notation and use $\varphi(U, R)$ to denote the number of nodes active at convergence when the initial active set is U and the randomness used by the process is R. Note that this is no longer random. By representing the randomness explicitly, we can easily express concepts such as the distribution over $\varphi(U \cup \{w\}, R)$ given that $\varphi(U, R) = Z$ for some fixed set Z. This is simply the distribution over $\varphi(U \cup \{w\}, R)$ that arises when we choose R uniformly at random from all possible values R such that $\varphi(U, R) = Z$.

With this notation in place, the first step of the proof is to show that the distribution over the set $\varphi(U \cup \{w\}, R) \setminus \varphi(U, R)$ conditioned on $\varphi(U, R) = Z$ is the same as the distribution over the set $\varphi^Z(\{w\})$. It is easiest to argue this point by considering the threshold view of the process. In particular, the argument is based on the simple observation that delays of arbitrary length can be introduced into the threshold model with no effect on the final set of active nodes. In other words, because we are restricting our attention to models in which the probability that a node is activated after a set S of his neighbors have tried to activate him is independent of the order in which they try, it is not important whether a node that is activated by a neighbor after time t becomes active right at time $t + 1$ or at a later time t'. This notion is formalized in Lemma 2 of Kempe et al. [37].

Because of this delay-invariance, we can think about w in the original process as a special node that, if activated, does not attempt to activate his neighbors until all other nodes have converged. At this point in time, w is the only "contagious" node. Since we have conditioned on the event that $\varphi(U, R) = Z$ and assumed that $w \notin Z$, it must be the case that at this point in time, precisely the set of nodes Z are active. It is easy to argue that the set of nodes activated from this point on (i.e., $\varphi(U \cup \{w\}, R) \setminus \varphi(U, R)$) is precisely the set of nodes that would be activated in the residual process if the same random threshold values were used for the remainder of the process. Since conditioning on $\varphi(U, R) = Z$ does not limit the randomness used in the remainder of the decisions, the distributions are the same. This gives us that

$$\begin{aligned} \mathrm{E}\Big[\big|\varphi(U \cup \{w\}, R) \setminus \varphi(U, R)\big|\Big] &= \sum_Z \mathrm{P}(\varphi(U, R) = Z)\mathrm{E}\Big[\big|\varphi(U \cup \{w\}, R) \setminus \varphi(U, R)\big|\Big|\varphi(U, R) = Z\Big] \\ &= \sum_Z \mathrm{P}(\varphi(U, R) = Z)\sigma^Z(\{w\}) \ , \end{aligned} \tag{2}$$

where all expectations and probabilities are over the choice of R.

Now, the decreasing cascade condition implies that for any sets Z and Z' with $Z \subseteq Z'$, for all neighboring nodes v and u and sets $S \in \mathcal{N}(v) \setminus \{u\}$,

$$p^{Z}_{v,u}(S) = p_{v,u}(S \cup (Z \cap \mathcal{N}(v))) \geq p_{v,u}(S \cup (Z' \cap \mathcal{N}(v))) = p^{Z'}_{v,u}(S) \ .$$

Intuitively, this implies that if $Z \subseteq Z'$, then

$$\sigma^{Z}(\{w\}) \geq \sigma^{Z'}(\{w\}) \ . \tag{3}$$

In other words, since the ground set of nodes $V \setminus Z$ is larger than the set $V \setminus Z'$, and the individual activation probabilities are higher in the residual process corresponding the Z, the set of nodes that are activated at convergence should also be larger. This statement can be proved formally by applying Lemma 3 of Kempe et al. [37].

From Equations 2 and 3, we have that for any sets U and U' such that $U \subseteq U' \subset V$, and any $w \in V$,

$$\begin{aligned}
\sigma(U \cup \{w\}) - \sigma(U) &= \mathrm{E}\Big[\big|\varphi(U \cup \{w\}, R)\big|\Big] - \mathrm{E}\Big[\big|\varphi(U, R)\big|\Big] \\
&= \mathrm{E}\Big[\big|\varphi(U \cup \{w\}, R) \setminus \varphi(U, R)\big|\Big] \\
&= \sum_{Z} \mathrm{P}(\varphi(U, R) = Z)\sigma^{Z}(\{w\}) \\
&= \sum_{Z} \sum_{Z' : Z \subseteq Z'} \mathrm{P}(\varphi(U, R) = Z, \varphi(U', R) = Z')\sigma^{Z}(\{w\}) \qquad (4) \\
&\geq \sum_{Z} \sum_{Z' : Z \subseteq Z'} \mathrm{P}(\varphi(U, R) = Z, \varphi(U', R) = Z')\sigma^{Z'}(\{w\}) \\
&= \sum_{Z'} \mathrm{P}(\varphi(U', R) = Z')\sigma^{Z'}(\{w\}) \\
&= \sigma(U' \cup \{w\}) - \sigma(U') \ .
\end{aligned}$$

Here we have again explicitly introduced the randomization R to make it clear that in Equation 4, $\mathrm{P}(\varphi(U, R) = Z, \varphi(U', R) = Z')$ refers to the joint probability that, using the *same underlying randomization*, the initial cascade process would produce active nodes Z starting with the active set U and would produce active nodes Z' starting with the active set U'. Because the decreasing cascade is monotonic, this probability would be 0 if it were not the case that $Z \subseteq Z'$.

Thus the influence function for the decreasing cascade model is both monotonic and submodular, and the greedy algorithm can be used to find the initial active set that approximately maximizes influence whenever this natural diminishing returns property is satisfied.

3.1.4 Extensions

In the previous section, we illustrated that the influence function for the decreasing cascade model is submodular, implying that approximate k-best influence maximization is possible using a simple greedy algorithm. In fact, Mossel and Roch [49] recently proved a conjecture of Kempe et al. that states that the influence function is always submodular for a much broader class of models. Specifically, they show that the influence function is submodular for the threshold model whenever the following three assumptions hold for all $v \in V$:

1. The threshold θ_v is chosen uniformly at random in $(0, 1)$, independent of the thresholds of other nodes.
2. The activation function f_v is monotone increasing.
3. The activation function f_v is submodular.

When translated to the threshold model via Equation 1, the decreasing cascade model meets these three requirements. Thus the results of the previous section can be obtained by applying this result directly.

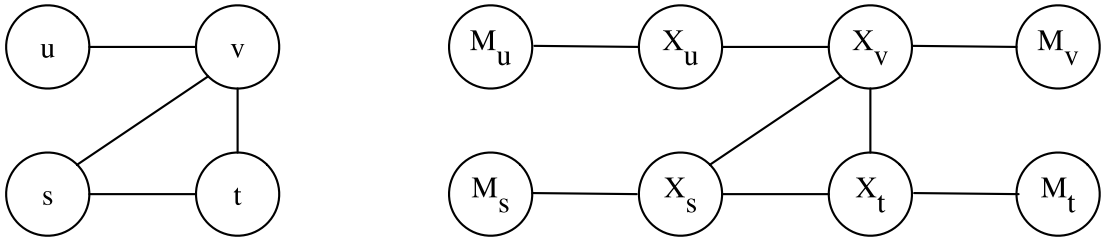

Figure 2: On the left is a simple social network. On the right is the corresponding Markov network. For each node in the social network, two nodes have been created in the Markov network.

Chen [9] studied complexity issues involved with the related problem of finding the smallest set $\mathcal{A}$ such that setting $\mathcal{A}_1 = \mathcal{A}$ ensures that the *entire* network becomes active in a deterministic version of the threshold model. This problem is generally hard to approximate within a polylogarithmic factor, but exact solutions can be computed efficiently in some special cases, such as when the underlying graph is a tree.

Even-Dar and Shapira [25] examined the influence maximization problem in the related voter model. In the voter model, at each time step, each node chooses one of his neighbors at random and adopts the current action of this neighbor. This model, while similar to the threshold model in some respects, is more suitable in scenarios in which progressiveness is undesirable and has the nice property that it is guaranteed to converge to a consensus (either everyone chooses the new action A or everyone chooses the incumbent action B) with probability 1. Even-Dar and Shapira show that when the cost of marketing to each individual in the network is the same, the obvious heuristic solution of marketing to those individuals with the highest degree is in fact optimal in this setting, and give a fully polynomial time approximation scheme that works when this is not the case.

Many other extensions of these models are possible. For example, one might consider allowing more complex marketing actions in place of the binary "market" or "do not market" options. In their earlier work, Kempe et al. [36] show that efficient approximation results can be achieved using the same greedy algorithm in such a setting when the effects of marketing satisfy a natural diminishing returns property.

3.2 Maximizing Influence Using Markov Networks

We now turn to the somewhat different variant of the influence maximization problem studied by Domingos and Richardson [21, 53]. For clarity of exposition, we focus on a simplified version of their model. This simplification is rich enough to illustrate the main ideas and intuition behind the model, without weighing down the explanation with superfluous details. Extensions of this basic model are discussed in Section 3.2.2.

Consider a social network represented by an undirected graph $\mathcal{G} = (V, E)$. For each $v \in V$, we define two corresponding Boolean random variables. The variable X_v corresponds to whether or not v eventually adopts the product; it takes on the value 1 if v becomes active and 0 otherwise. The variable M_v represents the extent to which v is marketed to. For now, this can be thought of as a binary variable that is 1 if v is targeted for advertising, and 0 otherwise. We use the shorthand $\mathbf{N}_v$ to denote the set of variables X_u for all $u \in \mathcal{N}(v)$. We denote by $\mathbf{X}$ the set of variables X_v for all $v \in V$, and by $\mathbf{M}$ the set of all M_v.

Together, these sets of variables naturally form a Markov network. In particular, for all $u, v \in V$, an edge exists between X_u and X_v in the Markov network if and only if $(u, v) \in E$. Additionally, for each v, there is an edge between X_v and M_v, representing the effect of direct marketing on each individual's decision to purchase the product. An example of the Markov network corresponding to a simple social network is shown in Figure 2.

Note that unlike the other models we have examined, there is no explicit notion of dynamics or time in this model. The model tells us only the probability that each node is activated at some point during the process and says nothing about the particular order in which the activations occur.

The goal of Domingos and Richardson can be expressed as setting $\mathbf{M}$ to maximize the global expected

lift in profit. As mentioned earlier, the expected lift in profit (denoted ELP) is simply the difference between the expected profit obtained by employing a marketing strategy $\mathbf{M}$ and the expected profit obtained using no marketing. Let $\mathbf{M}_0$ represent the strategy in which no marketing is employed. Formally,

$$ELP(\mathbf{M}) = \sum_{v \in V} \mathrm{P}(X_v = 1|\mathbf{M}) - \sum_{v \in V} \mathrm{P}(X_v = 1|\mathbf{M}_0) - c|\mathbf{M}|, \tag{5}$$

where c is the cost of marketing to a single individual and $|\mathbf{M}|$ is the number of v such that $M_v = 1$. We assume that c is normalized so that the profit gained by a single node adopting the product is 1.

In their earlier work, Domingos and Richardson [21] discuss various heuristics and techniques that can be used to approximately calculate the expected lift in profit. Unfortunately, there are a number of flaws in this first attempt that make the applicability of such an approximation questionable.

First, in order to calculate $\mathrm{P}(X_v|\mathbf{M})$ for Equation 5, Domingos and Richardson employ a discrete relaxation labeling technique as in Pelkowitz [51]. In this context, relaxation labeling refers to a class of iterative algorithms in which at each time step, a label (in this case, the current estimate of $\mathrm{P}(X_v = 1|\mathbf{M})$) is assigned to each vertex based on the current labels of the other vertices. This process is repeated until a fixed point is reached. In order to get an expression for $\mathrm{P}(X_v = 1|\mathbf{M})$ that is in the right form to apply such an iterative technique, Domingos and Richardson start with the equation

$$\mathrm{P}(X_v|\mathbf{M}) = \sum_{\mathbf{N}_v} \mathrm{P}(X_v|\mathbf{N}_v, \mathbf{M})\mathrm{P}(\mathbf{N}_v|\mathbf{M}) \ ,$$

where the sum over $\mathbf{N}_v$ is meant to represent a sum over all possible assignments to the variables X_u for each $u \in \mathcal{N}(v)$. Still following Pelkowitz [51], they replace the term $\mathrm{P}(\mathbf{N}_v|\mathbf{M})$ in this equation with its maximum entropy estimate given the marginals $\mathrm{P}(X_u|\mathbf{M})$, essentially assuming

$$\mathrm{P}(\mathbf{N}_v|\mathbf{M}) \approx \prod_{u \in \mathcal{N}(v)} \mathrm{P}(X_u|\mathbf{M}) \ ,$$

where X_u is the assignment to node u specified by $\mathbf{N}_v$. In other words, in order to calculate $\mathrm{P}(X_v|\mathbf{M})$, they roughly assume that the variables X_u for all $u \in \mathcal{N}(v)$ are independent given $\mathbf{M}$. In practice, this may be a very bad assumption. It is, in essence, assuming that any network effects are so minimal that they can be ignored, which is against the very spirit of the work. Additionally, in practice, it is often the case that two nodes with a common neighbor are *more likely* to be neighbors with each other than two nodes without a common neighbor [29, 12], so it is likely that the set of neighbors of a node have especially great influence on each other.

Second, they assume that for all $v \in V$, it is possible to accurately estimate $\mathrm{P}(X_v|M_v)$ and $\mathrm{P}(M_v|X_v)$ from data, arguing that $\mathrm{P}(M_v|X_v)$ can be estimated via "a data collection phase in which users to market to are selected at random and their responses are recorded (both when being marketed to and not)." It is not clear that devoting time and budget to such a data collection phase is feasible in general. Furthermore, even if it were possible to collect data in this way, it is not clear how this data can be used to obtain an estimate of $\mathrm{P}(M_v|X_v)$ that will be accurate when the decision of whether or not to market to each individual is no longer made uniformly at random. In general, the entire discussion of estimating $\mathrm{P}(M_v|X_v)$ is somewhat odd given that the values of M_v will be fixed to arbitrary values when marketing is employed. A discussion of $\mathrm{P}(X_v|M_v)$ would be more natural.

Third, even if these quantities are known and the expected lift in profit can be calculated exactly, the problem of *maximizing* the expected lift in profit is still intractable, even for binary marketing actions, so another level of approximation is needed. Domingos and Richardson suggest using a greedy search or hill-climbing search, but neither is guaranteed to find a strategy that is close to optimal.

3.2.1 The Benefit of Linearity

In follow-up work, Richardson and Domingos [53] get around the first and third problems by imposing linearity assumptions on their model. In particular, they assume that for all v,

$$\mathrm{P}(X_v|\mathbf{N}_v, M_v) = \beta_v \mathrm{P}_0(X_v|M_v) + (1 - \beta_v)\mathrm{P}_N(X_v|\mathbf{N}_v)$$

where $\mathrm{P}_0(X_v|M_v)$ is a network-free function representing how likely it is that v would adopt the technology on his own, $\mathrm{P}_N(X_v|\mathbf{N}_v)$ is a function representing the influence of the neighbors of v, and $\beta_v \in [0,1]$ is a parameter measuring how much v depends on his neighbors' influence. They make the further assumption that P_N is linear. That is,

$$\mathrm{P}_N(X_v = 1|\mathbf{N}_v) = \sum_{u \in \mathcal{N}v} \omega_{v,u} X_u$$

where $\omega_{v,u} \geq 0$ for all v and u, and $\sum_{u \in \mathcal{N}v} \omega_{v,u} = 1$. This linear model is similar in spirit to the linear threshold model, but the dynamic aspects of the activation process are now implicit.

As before, if we are interested in maximizing the expected lift in profit, it is necessary to be able to calculate $\mathrm{P}(X_v|\mathbf{M})$. By marginalization,

$$\begin{aligned}
\mathrm{P}(X_v = 1|\mathbf{M}) &= \sum_{\mathbf{N}_v} \mathrm{P}(X_v = 1, \mathbf{N}_v|\mathbf{M}) = \sum_{\mathbf{N}_v} \mathrm{P}(X_v = 1|\mathbf{N}_v, M_v)\mathrm{P}(\mathbf{N}_v|\mathbf{M}) \\
&= \sum_{\mathbf{N}_v} \left(\beta_v \mathrm{P}_0(X_v = 1|M_v) + (1-\beta_v) \sum_{u \in \mathcal{N}(v)} \omega_{v,u} X_u \right) \mathrm{P}(\mathbf{N}_v|\mathbf{M}) \\
&= \beta_v \mathrm{P}_0(X_v = 1|M_v) + (1-\beta_v) \sum_{\mathbf{N}_v} \sum_{u \in \mathcal{N}(v)} \omega_{v,u} X_u \mathrm{P}(\mathbf{N}_v|\mathbf{M}) \\
&= \beta_v \mathrm{P}_0(X_v = 1|M_v) + (1-\beta_v) \sum_{u \in \mathcal{N}(v)} \omega_{v,u} \mathrm{P}(X_u = 1|\mathbf{M}) \ . \qquad (6)
\end{aligned}$$

The last line follows from the fact that since X_u is either 0 or 1 for all u,

$$\sum_{\mathbf{N}_v} X_u \mathrm{P}(\mathbf{N}_v|\mathbf{M}) = \mathrm{P}(X_u = 1|\mathbf{M}) \ .$$

Now for each v, we have an expression for $\mathrm{P}(X_v = 1|\mathbf{M})$ in terms of the values of $\mathrm{P}(X_u = 1|\mathbf{M})$ for all $u \in \mathcal{N}(v)$. Furthermore, these expressions were derived without requiring any unrealistic independence assumptions like the assumptions necessary for the more general model. Assuming that good approximations are known for the self-reliance factors β_v and weights $\omega_{v,u}$ and for $\mathrm{P}_0(X_v = 1|M_v)$, this gives us $|V|$ linear equations with $|V|$ unknown variables, so it is possible to efficiently solve the equations. In practice, such a system can be solved by iteratively updating the values of $\mathrm{P}(X_v = 1|\mathbf{M})$ for each v. This iterative relaxation labeling process is guaranteed to converge as long as the initial estimates are reasonably good [51].

Unfortunately, we are left with the problem of estimating the quantities β_v, $\omega_{v,u}$, and $\mathrm{P}_0(X_v = 1|M_v)$ from data. Richardson and Domingos [53] claim that estimating β_v and $\omega_{v,u}$ is easy given the resources of a typical marketing research department, though it is not clear that this is generally true in practice. When there is insufficient data to obtain good estimates, they suggest using "a combination of prior knowledge and any demographic information available." Again, it is not clear that such approximations are sufficient, although they are reasonable choices when no other information is available.

Estimating $\mathrm{P}_0(X_v|M_v)$ is trickier, and Richardson and Domingos do not give a general method for doing so. To obtain their experimental results, they set

$$\mathrm{P}_0(X_v = 1|M_v = 1) = \min\{\alpha \mathrm{P}_0(X_v = 1|M_v = 0), 1\}$$

for some $\alpha > 0$, adding the additional constraint that $\mathrm{P}(M_v = 1) = \mathrm{P}(M_v = 0)$. The reasoning behind the addition of this counterintuitive constraint is not explained.

Assuming that accurate estimates of the parameters for each individual are known, it is possible to efficiently maximize $ELP(\mathbf{M})$. At a high level, this is because the linearity of the model causes the network effect of marketing to one individual to be independent of the network effect of marketing to others. As a result, the decision of whether or not to market to one particular node can be made independently of all other marketing decisions. Richardson and Domingos [53] provide a full analysis of this phenomenon in a

general setting in which marketing strategies may be continuous. In their analysis, they separate out the so-called "intrinsic value" and the "network value" of each node and analyze the total increase in purchasing probability in the network that results from a small unit change in marketing to a particular node. For intuition, here we instead stick with the assumption that marketing actions are binary, and examine the effect of marketing to one individual.

Let $\mathbf{M}$ be an arbitrary marketing strategy in which $M_s = 0$ for some $s \in V$, and let $\mathbf{M}'$ be the same strategy with M_s set to 1. From Equation 5,

$$ELP(\mathbf{M}') - ELP(\mathbf{M}) = \sum_{v \in V} \big(\mathrm{P}(X_v = 1|\mathbf{M}') - \mathrm{P}(X_v = 1|\mathbf{M})\big) - c \ . \tag{7}$$

From Equation 6, for any $v \neq s$,

$$\mathrm{P}(X_v = 1|\mathbf{M}') - \mathrm{P}(X_v = 1|\mathbf{M}) = (1 - \beta_v) \sum_{u \in \mathcal{N}(v)} \omega_{v,u} \big(\mathrm{P}(X_u = 1|\mathbf{M}') - \mathrm{P}(X_u = 1|\mathbf{M})\big) \ ,$$

and

$$\begin{aligned} \mathrm{P}(X_s = 1|\mathbf{M}') - \mathrm{P}(X_s = 1|\mathbf{M}) \quad &= \quad \beta_s \left(\mathrm{P}_0(X_s = 1|M_s') - \mathrm{P}_0(X_s = 1|M_s)\right) \\ &\quad + (1 - \beta_s) \sum_{u \in \mathcal{N}(s)} \omega_{s,u} \big(\mathrm{P}(X_u = 1|\mathbf{M}') - \mathrm{P}(X_u = 1|\mathbf{M})\big) \ . \end{aligned}$$

As before, we have an expression for a particular quantity for each node v (namely, $\mathrm{P}(X_v = 1|\mathbf{M}') - \mathrm{P}(X_v = 1|\mathbf{M})$) in terms of the same quantity for the other nodes. Thus once again we have $|V|$ linear equations and $|V|$ unknowns, and can, in practice, solve the system of equations iteratively to find the value of $\mathrm{P}(X_v = 1|\mathbf{M}') - \mathrm{P}(X_v = 1|\mathbf{M})$ for each v. Note that these differences are identical for *all strategies* $\mathbf{M}$ with $M_s = 0$. In other words, the difference in expected lift in profit marketing to v versus not marketing to v does not depend on the other nodes that are targeted. Equation 7 thus tells us whether or not it is worth marketing to node s, independent of all other marketing actions.

3.2.2 Extensions of the Basic Model

As mentioned before, the model discussed in the previous section is a simplification of the full model of Richardson and Domingos [53]. Richardson and Domingos introduce a variety of extra parameters and variables into the model. First, they introduce a vector of attributes $\mathbf{Y}$ describing the product that is marketed. The distribution over X_v then depends directly on $\mathbf{Y}$ in addition to $\mathbf{N}_v$ and M_v. In their empirical experiments, they either use $\mathbf{Y}$ to coarsely divide the products into different groups [21] or ignore it entirely and focus only on a group of products with the same attributes [53].

Additionally, they allow for the possibility that the revenue generated from a sale to node v varies depending on the marketing actions used. This could be used, for example, to model a situation in which the marketing action consists of giving each targeted node a coupon or discount. However, in order to obtain a linear solution, they are later forced to backtrack and assume again that the revenue generated from each sale is a constant value. (Interestingly, they justify this assumption by stating that it is "safe" because "typically only a small portion of the network will be marketed to," which appears to blatantly contradict the assumption mentioned above that $\mathrm{P}(M_v = 1) = \mathrm{P}(M_v = 0)$ for all v.)

3.3 Discussion

As we have seen, although Kempe et al. and Richardson and Domingos are interested in roughly the same problem, they approach it from very different angles. To start, the two groups phrase the goal of influence maximization in slightly different ways; Richardson and Domingos attempt to maximize expected lift in profit, while Kempe et al. attempt to find the best set of k individuals to target for a fixed value of k. On the surface, the goal of maximizing the expected lift in profit appears more flexible and useful, given

that profit is what matters to a marketing company in the end. If marketing to more than k individuals is profitable, then intuitively the company should market to more than k individuals. However, in practice, it is unlikely that any real-world marketing department would be given access to arbitrary marketing resources. Indeed, marketing departments generally have fixed budgets, and thus an algorithm for finding the k best individuals to target is in some ways more practical.

Both models require that the individual parameters for each node in the network are known to the marketing firm. Kempe et al. assume knowledge of the individual threshold functions f_v or cascade probabilities $p_{v,u}$, while Richardson and Domingos assume knowledge of the self-reliance factor β_v and weights $\omega_{v,u}$, as well as $\mathrm{P}_0(X_v|M_v)$. Kempe et al. do not address the issue of how these parameters might be learned. Richardson and Domingos suggest heuristic methods of approximating these parameters, but provide is no guarantee that these heuristic methods are successful. Neither paper discusses the effect of using inexact estimates of these parameters on the outcome.

Indeed, in all discussions of empirical results, the potential harm of estimating parameters from data is ignored. Domingos and Richardson provide extensive experiments based loosely on data provided from the collaborative filtering database EachMovie[2] and the knowledge sharing site Epinions,[3] but in both cases parameters are estimated somewhat arbitrarily. For example, in the Epinions experiment, they assume that $\beta_v = 0.5$ for all v, and assume that all neighbors u of v have equal weight $\omega_{v,u}$. Additionally, they make the aforementioned counterintuitive assumption that $\mathrm{P}(M_v = 1) = \mathrm{P}(M_v = 0)$ for all v. In both experiments, the real data is used only to provide a graph structure and in some cases to obtain loose estimates of the necessary probabilities. Their hypothetical comparisons of various methods assume that the estimated probabilities are correct. The same is true of the empirical results from of Kempe et al. [36], who state this weakness up front. This assumption could potentially prevent either set of results from being applicable in practice, and should be examined more carefully.

The linear version of the Richardson and Domingos [53] model has the advantage of being efficiently solvable. However, since it is again unclear how bad of an approximation might result from the use of estimated parameters, it is not obvious that the results are better than those obtained using the greedy approximation algorithm of Kempe et al., which is also efficient.

Despite the potential drawbacks mentioned here, the work surveyed in this section is important in that it has inspired a new active line of research. Domingos and Richardson posed a question that had never before been asked in the computer science community, and their work, in addition to the early work of Kempe et al., caused a flourish of interest in viral marketing and diffusion properties of networks [40, 31, 43, 44, 38, 46, 49, 25, 5, 9]. The body of research on network diffusion in the computer science literature continues to grow today because of these initial investigations into influence maximization and related problems.

4 Graph Properties Linked to Contagion

The previous section examined the problem of influence maximization, surveying methods for choosing the optimal set of nodes to market to in order to ensure that a trend spreads across a network. In this section, we turn to the related but fundamentally different study of the properties of infinite graphs that are linked to the spread of trends.

Morris [48] was one of the first to examine this question in depth. By studying best response dynamics on local interaction games, he aimed to uncover the diffusion properties of different classes of infinite graphs. Specifically, his work attempts to characterize the set of graphs and values of the local interaction game parameter q (see Section 2.1) for which there exists a finite set of players $\mathcal{A}_1$ such that if only this set choose action A at time 0, and all other players choose the incumbent action B, action A is eventually chosen everywhere in the network. He does not care *which* or even *how many* nodes are active initially, as long as it is only a finite set. Additionally, he is concerned only with whether or not the new trend or technology spreads to the *entire* network. In this sense, his work is more focused on the analysis of a particular set

[2]The EachMovie data set was available at `http://www.research.compaq.com/src/eachmovie` when the experiments were initially run. However, it is no longer publicly available from HP/Compaq.

[3]This data set was scraped from `http://www.epinions.com`.

of properties of graphs as opposed to the algorithmic aspects of viral marketing with which Domingos and Richardson and Kempe et al. were concerned.

4.1 A Characterization of Diffusion

Consider an infinite network of players engaging in a local interaction game. To start, it is necessary to define what is meant by a technology spreading to the entire infinite network.

Definition 2 *We say that A is* epidemic on graph $\mathcal{G}$ with parameter q *if there exists a finite set $\mathcal{A}_1$ such that*

$$\bigcup_{t \geq 1} \mathcal{A}_t = V$$

where for every $t \geq 1$, $\mathcal{A}_t$ is the set of players choosing action A at time t assuming that players each play a best response to their neighbors' actions at time $t-1$ under parameter q. We define the contagion threshold *τ of a graph $\mathcal{G}$ as the maximum value of q for which A is epidemic on $\mathcal{G}$ with parameter q.*

Note that this setting is not explicitly defined to be progressive. In other words, there is nothing explicitly preventing nodes currently playing action A from switching back to action B if B is a best response. However, if it is assumed that the initial set $\mathcal{A}_1$ stick with action A, then it can easily be shown that no node ever switches from action A to action B. Furthermore, as defined above, the contagion threshold is actually *the same* whether we allow the set $\mathcal{A}_1$ of nodes who initially begin choosing A to update their actions based on best response dynamics or whether we enforce that their actions remain fixed.

It is perhaps easiest to understand this definition through a simple example. Consider an infinite population interacting along a line. Here we can index each vertex $v \in V$ by an integer i. Then $\mathcal{N}(i) = \{i-1, i+1\}$. Suppose that $q \leq 1/2$. In other words, A is at least as good a technology as B, or better. In this case, it is a best response to choose action A whenever at least one neighbor chooses action A. Suppose a single node i chooses action A at time 1. At time 2, the two neighbors of this node, $i-1$ and $i+1$ both switch to A. At time 3, A spreads to $i-2$ and $i+2$. At each subsequent time t, nodes $i-(t-1)$ and $i+(t-1)$ choose action A, and A eventually spreads throughout the entire network.

If, on the other hand, $q > 1/2$, then no node ever has incentive to switch to action A unless *both* of his neighbors choose action A. This implies that no finite group of nodes can cause a widespread epidemic of action A. For the line graph, then, the contagion threshold is $1/2$.

In fact, it can be shown that the contagion threshold is never more than $1/2$ for any graph $\mathcal{G}$. Intuitively, this means that it is never possible in this model for a weaker technology to overcome a stronger incumbent technology. However, contagion thresholds strictly less than $1/2$ are possible. Morris gives a detailed analysis of the properties of a graph that impact the contagion threshold.

One important and intuitively simple property is the *cohesion* of sets of nodes in the graph. At a high level, cohesion is a measure of how "cliquey" a social network is, comparing the frequency of links within a group of nodes to the frequency of links between members of the group and outside nodes. More formally, the cohesion of a set $S \subset V$ is the largest value of p such that every node in S has a proportion p or more of his neighbors in S. For example, on the line graph, every finite, non-empty set has cohesion at most $1/2$, because for any finite set there must exist at least one node on the boundary with a neighbor outside the set.

To see how cohesion affects the contagion properties of a graph, consider a graph that contains a set S with cohesion p, and suppose that every node in this set initially chooses action B. Each node switches to action A only if a fraction q or more of his neighbors switch to A. If $p > q$, then no node in S ever switches to A. In this way, high levels of cohesion in a graph can prevent the spread of trends.

A second relevant property is what Morris calls *neighbor growth*. Let $\mathcal{N}_n(v)$ be the set of nodes within n hops of v on the graph, with $\mathcal{N}_1(v) = \mathcal{N}(v)$. Formally, the graph has low neighbor growth if for all nodes v, $|\mathcal{N}_n(v)| = o(\gamma^n)$ for some $\gamma > 1$; that is, $|\mathcal{N}_n(v)|/\gamma^n \to 0$ as $n \to \infty$. This property is closely related to the notion of "weak" versus "strong" social links introduced by Granovetter [29]. Strong social links, such as ties between chose friends, are frequently transitive; if v and u are both close friends with w, then they

are more likely to be friends with each other than two nodes chosen at random from the population. Weak social links are less likely to exhibit this property, and are more likely to result in graphs with low neighbor growth. Empirical studies confirm that $|\mathcal{N}_n(v)|$ tends to grow more slowly for graphs in which the edges represent important relations than for graphs describing weaker ties [52].

The effect of neighbor growth on contagion is less obvious and more technically involved than the effect of cohesion. Essentially, low neighbor growth, in conjunction with a technical "uniformity" property, can be used to show a lower bound on the contagion threshold, implying that it is possible for contagion to occur even when the value of q is relatively high (but still less than 1/2).

4.2 Introducing Compatible Technologies

Consider again the scenario in which a network of friends are choosing which online chat client to install. If an individual's neighbors are truly a mix of Yahoo! Messenger users and AOL Instant Messenger users, it might be in the interest of the individual not to choose a single client to install, but to in fact install both. This option would allow the individual to chat with *all* of his friends online, in exchange for the added hassle of maintaining two separate chat clients.

Immorlica et al. [32] introduce a new model, which we refer to as *compatible contagion games*. Compatible contagion games are a natural extension of local interaction games to the scenario in which players may choose to adopt either product A or product B as before, or to instead adopt both products A and B simultaneously at some additional cost. This cost may be high or low, depending on the inherent compatibility between products A and B. For example, it might require very little effort for a user to install two instant messenger clients in order to communicate with all of his friends. On the other hand, the additional price of service and hassle of carrying around multiple phones would generally prohibit users from signing up for cell phone service from two competing companies just to gain the free mobile-to-mobile minutes from each.

Immorlica et al. again assume that the underlying graph is infinite. They make the additional, quite restrictive assumption that every node on the graph has degree Δ for some fixed Δ. They denote by c the total penalty a node must pay in order to simultaneously adopt both product A and product B. Because the graph is regular, this cost can be viewed as a per-neighbor cost of $r = c/\Delta$. Each set of neighbors can then be viewed as playing the modified coordination game in Table 2. As before, the total payoff to a player is the sum of his payoffs in the coordination games played with each neighbor.

	A	B	AB
A	$1-q,\ 1-q$	$0,\ 0$	$1-q,\ 1-q-r$
B	$0,\ 0$	$q,\ q$	$q,\ q-r$
AB	$1-q-r,\ 1-q$	$q-r,\ q$	$\max\{q,1-q\}-r,\ \max\{q,1-q\}-r$

Table 2: A modified coordination game in which each player may choose to adopt both technologies A and B at a price.

The analog of the contagion threshold in the compatible contagion game setting is a two-dimensional *epidemic region* $\Omega(\mathcal{G})$ consisting of all (q, r) pairs for which A can become epidemic on $\mathcal{G}$. The epidemic region can be surprisingly complex, even for very simple graphs, as the following example shows.

4.2.1 Analysis of the Line Graph

Recall the simple line graph introduced in Section 4.1 in which each vertex $v \in V$ is indexed by an integer i and $\mathcal{N}(i) = \{i-1, i+1\}$. This graph is regular with $\Delta = 2$, and thus can be analyzed in the compatible contagion game setting. Assume that $\mathcal{A}_1 = \{0\}$. That is, node 0 initially chooses the new action A and all other nodes initially choose the incumbent action B. We are interested in finding all settings of the parameters q and r for which action A becomes epidemic on the graph.

Consider the decision that node 1 (and symmetrically, node -1) faces after the first time step. One of his neighbors is choosing action A, and one is choosing action B. By examining Table 2, we see that he receives

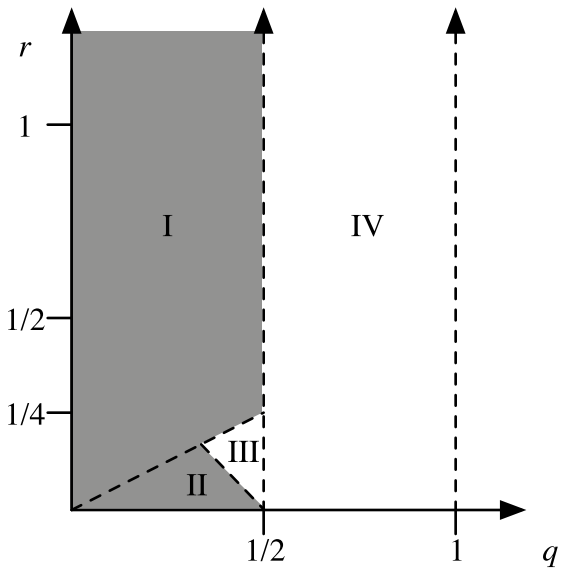

Figure 3: Illustration of the line graph example. The epidemic region is displayed in gray. When the parameters fall in region I, nodes with one neighbor choosing A and one neighbor choosing B switch to A, causing an immediate epidemic. When the parameters fall in region II or III, nodes with one neighbor choosing A and one neighbor choosing B switch to AB. In region II, nodes with one AB neighbor also switch to AB, and eventually to A, while in region III, nodes with one AB neighbor and one B neighbor stick with action B. Region IV represents the area of parameter space in which B is a better technology. In this region, nodes switch to A only if both neighbors are choosing A, and A never becomes epidemic.

a payoff of $1-q$ for choosing action A, q for choosing action B, or $(1-q-r)+(q-r) = 1-2r$ for choosing AB. Thus A is a best response if $q \leq 1/2$ and $q \leq 2r$. If this is the case, then nodes 1 and -1 both choose A at time 2, causing nodes 2 and -2 to choose A at time 3, and so on, and A becomes epidemic.

Suppose this is not the case. If $2r < q \leq 1-2r$, then AB is instead the best response for nodes 1 and -1 after time 1, and both choose AB at time 2. At this point, node 2 (and symmetrically, node -2) has one neighbor choosing AB and one neighbor choosing B. By referring again to Table 2, we see that she receives a payoff of $1-q$ for choosing A, $2q$ for choosing B, or $q + \max\{q, 1-q\} - 2r = \max\{2q, 1\} - 2r$ for choosing AB. Some simple algebra shows that AB is her best response if and only if $q \geq 2r$ and $q + r \leq 1/2$. If this is the case, then AB continues to spread to nodes 3 and -3, then to nodes 4 and -4, and so on.

As AB continues to spread, it soon becomes the case that all of these nodes have *two* neighbors choosing AB. Once both of a node's neighbors are choosing AB, it is no longer beneficial for the node itself to choose AB and pay the compatibility cost. At this point, the node can instead switch to just action A as long as $q \leq 1/2$, and action A again becomes epidemic.

4.2.2 Interpretation of the Epidemic Region

Figure 3 shows the full epidemic region of the line graph. Notice that this region is not convex. The irregular shape of the region, which turns out to be quite common, has an interesting economic interpretation.

When the cost r of simultaneously adopting both products A and B is relatively large, as in the example of the cell phone provider or region I of Figure 3, nodes that have a mix of neighbors choosing A and neighbors choosing B have incentive to choose only action A. This causes action A to become epidemic on the network, even if $\mathcal{A}_1$ consists of only a single node.

When the cost of simultaneously adopting both products is relatively small, as in the example of the competing chat clients, and product A is sufficiently better quality than product B, as in region II of Figure 3,

nodes with a mix of neighbors choosing A and B choose to adopt both, and the action AB quickly spreads throughout the network. However, once the neighbors of a node are choosing AB, the node has incentive to switch to action A alone, which allows him to receive all of the benefits of compatibility without paying the extra cost. Thus there is a second wave in which action A again becomes epidemic.

On the other hand, consider the scenario in which either the cost of adopting both products is in between these extremes, or the cost of adopting both is low but the difference in quality is small, as in region III of Figure 3. Nodes with a mix of A and B neighbors still prefer to choose AB since the cost of adopting both products is not too large. However, the cost of adopting both is sufficiently prohibitive that once a node has chosen AB, his A-choosing neighbors have no incentive to deviate from A, and his B-choosing neighbors have no incentive to deviate from B, causing the spread to halt.

In this sense, the producers of the new product A have incentive to make A either very compatible or very incompatible with B, while the producers of the incumbent product B would prefer a balance.

4.2.3 A More General Characterization

Immorlica et al. provide a characterization of the epidemic region analogous to the original characterization of the contagion threshold. To start, they show that compatible contagion games satisfy a property very similar to progressiveness. In particular, if the initial set of active nodes $\mathcal{A}_1$ are assumed to choose action A at every time step, then no node ever switches from action A to action B or AB. Furthermore, no node ever switches from action AB to action B. This property implies that the actions of the nodes eventually converge, since loops in behavior are not possible.

With this result in place, they prove that the existence of a particular "blocking structure" in a graph $\mathcal{G}$ is sufficient to fully determine whether or not action A can become epidemic on $\mathcal{G}$ with parameters q and r. The blocking structure is defined as follows, where the notation $deg_S(v)$ is used to denote the number of neighbors that node v has in the set S.

Definition 3 *Consider a compatible contagion game $(\mathcal{G}, q, r)$, where $\mathcal{G} = (V, E)$ is an infinite, undirected, Δ-regular graph. A pair (S, T) of disjoint, non-empty subsets of V is a* blocking structure *for this game if the following conditions hold.*

1. *For every $v \in S$, $deg_T(v) > \Delta r/q$.*
2. *For every $v \in T$, $(1-q)deg_T(v) + \min\{q, 1-q\}deg_S(v) > (1-q-r)\Delta$.*
3. *For every $v \in T$, $deg_T(v) + q \cdot deg_S(v) > (1-q)\Delta$.*

Although this blocking structure is quite complex, the intuition behind it is similar to the intuition behind the problems caused by cohesiveness in the model of Morris [48]; if each vertex in $S \cup T$ has a sufficient fraction of their neighbors in these two sets, then the group is hard to penetrate. In particular, suppose that such a blocking structure exists, and that none of the nodes in S or T are in the initial active set. It is possible to show that after any sequence of moves, every node in S chooses either action B or action AB, and every node in T chooses action B. This yields the following theorem, the proof of which is somewhat technical but mostly involves careful algebraic manipulation, and can be found in Section 4 of Immorlica et al. [32].

Theorem 4 *For every compatible contagion game $(\mathcal{G}, q, r)$, A cannot become epidemic if and only if every co-finite set of vertices of $\mathcal{G}$ contains a blocking structure.*

Additionally, Immorlica et al. analyze the region obtained by taking the union over all infinite, undirected, Δ-regular graphs $\mathcal{G}$ of the epidemic regions $\Omega(\mathcal{G})$. Just as in the local interaction game setting, it turns out that A can never become epidemic if $q > 1/2$. However, it is possible to show that there exist values of q and r with $q < 1/2$ such that for *every* compatible contagion game, A *cannot* become epidemic. This implies that the nonconvex shape of the epidemic region and the corresponding interpretation discussed in Section 4.2.2 are not limited to toy examples such as the line graph, but are part of a more general phenomenon.

4.3 Discussion

Pieces of the analysis provided in Immorlica et al. [32] are very suggestive. For example, as discussed in Section 4.2.2, there is a very interesting economic interpretation of the fact that the epidemic region is, in general, not convex. This interpretation could provide valuable insight to companies designing new technologies who want to maximize the chance that their technology will spread.

The main drawback of the analysis is the reliance on the assumption that the underlying graph $\mathcal{G}$ is Δ-regular. This assumption is critical to the results presented here, as it allows the cost c of using multiple technologies to be viewed as a per-neighbor cost of $r = c/\Delta$. However, in recent years, dozens if not hundreds of studies have been published illustrating that social networks are generally not regular, but instead typically exhibit power law distributions of degree [57, 2, 22, 8, 3]. Results that apply only to regular graphs are unlikely to be applicable to most real-life social networks.

Immorlica et al. do not discuss any attempts to get around this extremely unrealistic assumption. However, most of their analysis could be adapted to hold under the more direct assumption that the cost of using two technologies scales linearly with the number of neighbors, which could conceivably be a more reasonable assumption than regularity in some settings. Of course, the analysis would be applicable in an even wider variety of settings if the assumption could be removed altogether.

5 Conclusions

With the increasing popularity of product recommendation systems, social networking websites, and online chat programs, huge social networks are becoming ubiquitous. The models, algorithms, and analyses surveyed in this paper provide a framework for reasoning about the diffusion of innovations or trends on these large-scale networks. However, these techniques and models are only the beginning. Many open research questions remain.

As we have seen, inspired by the initial work of Domingos and Richardson [21, 53] and the novel questions this work raised, Kempe et al. [36, 37], Mossel and Roch [49], Even-Dar and Shapira [25], and others have developed a solid theoretical body of literature on the problem of influence maximization. An important next step is to work towards using this firm theoretical foundation to build solutions that are applicable in real viral marketing settings. The main challenge of this application is finding ways to learn or estimate the relevant modeling parameters for each individual in the social network without requiring too much expensive data or making unrealistic independence assumptions. Perhaps the original heuristic techniques suggested by Domingos and Richardson could prove useful here. In order to determine the real applicability of their approach, more interactive experiments, specifically those designed to test the accuracy of their approximation techniques, would be useful.

The more recent work of Immorlica et al. [32] also raises a number of interesting questions. As discussed, the theoretical characterization of the epidemic regions of graphs in the compatible contagion game would be significantly more interesting if it were extended to cover general, potentially nonregular graphs, especially if the resulting characterization shared the simplicity of the original. Such a characterization could provide useful insight for marketing firms, and could potentially influence other application areas as well, for example by providing insight on network structures that prevent the spread of computer viruses, or structures that allow information to be propagated quickly in an emergency. It is likely that the coming years will bring many more developments in this interesting line of research.

Acknowledgments

This report was written as part of my Written Preliminary Examination. I would like to thank Sanjeev Khanna, Ben Taskar, and Michael Kearns for serving on my WPE committee. I would also like to thank Sid Suri and Mohammad Mahdian for suggesting some starting points in the relevant literature. Most of all, I am sincerely grateful to Kuzman Ganchev, Jinsong Tan, and Jeff Vaughan for taking the time to read and comment on drafts of this document. I owe you one!

References

[1] Eytan Adar and Lada A. Adamic. Tracking information epidemics in blogspace. In *ACM International Conference on Web Intelligence*, 2005.

[2] Réka Albert and Albert-László Barabási. Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74:47–97, 2002.

[3] Noam Berger, Christian Borgs, Jennifer T. Chayes, and Amin Saberi. On the spread of viruses on the Internet. In *Proceedings of the 16th ACM-SIAM Symposium on Discrete Algorithm (SODA)*, 2005.

[4] Frances Stokes Berry and William D. Berry. State lottery adoptions as policy innovations: An event history analysis. *The American Political Science Review*, 84(2):395–415, 1990.

[5] Shishir Bharathi, David Kempe, and Mahyar Salek. Competitive influence maximization in social networks. In *Proceedings of the 3rd International Workshop On Internet And Network Economics (WINE)*, 2007.

[6] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.

[7] Lawrence E. Blume. The statistical mechanics of strategic interaction. *Games and Economic Behavior*, 5(3):387–424, 1993.

[8] Béla Bollobás and Oliver M. Riordan. Mathematical results on scale-free random graphs. In Stefan Bornholdt and Heinz Georg Schuster, editors, *Handbook of Graphs and Networks: From the Genome to the Internet*, pages 1–34. Wiley-VCH, 2003.

[9] Ning Chen. On the approximability of influence in social networks. In *Proceedings of the 19th ACM-SIAM Symposium on Discrete Algorithm (SODA)*, 2008.

[10] David Maxwell Chickering and David Heckerman. A decision-theoretic approach to targeted advertising. In *Proceedings of Sixteenth Conference on Uncertainty in Artificial Intelligence (UAI)*, 2000.

[11] Nicholas A. Christakis and James H. Fowler. The spread of obesity in a large social network over 32 years. *New England Journal of Medicine*, 357:370–379, 2007.

[12] Michael Suk-Young Chwe. Structure and strategy in collective action. *American Journal of Sociology*, 105:128–156, 1999.

[13] James S. Coleman, Elihu Katz, and Herbert Menzel. *Medical innovation: A Diffusion Study*. Bobbs-Merrill, Indianapolis, MN, 1966.

[14] Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein. *Introduction to Algorithms, Second Edition*. MIT Press and McGraw-Hill, 2001.

[15] Gerard Cornuejols, Marshall L. Fisher, and George L. Nemhauser. Location of bank accounts to optimize float: An analytic study of exact and approximate algorithms. *Management Science*, 23(8):789–810, 1977.

[16] Arthur De Vany. *Hollywood Economics: How Extreme Uncertainty Shapes the Film Industry*. Routledge, London, 2004.

[17] Arthur De Vany and Cassey Lee. Quality signals in information cascades and the dynamics of the distribution of motion picture box office revenues. *Journal of Economic Dynamics and Control*, 25: 593–614, 2001.

[18] Peter S. Dodds and Duncan J. Watts. Universal behavior in a generalized model of contagion. *Phyical Review Letters*, 92, 2004.

[19] Peter S. Dodds and Duncan J. Watts. A generalized model of social and biological contagion. *Journal of Theoretical Biology*, 232:587–604, 2005.

[20] Pedro Domingos. Mining social networks for viral marketing. *IEEE Intelligent Systems*, 20(1):80–82, 2005.

[21] Pedro Domingos and Matthew Richardson. Mining the network value of customers. In *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, 2001.

[22] Sergey N. Dorogovtsev and José Fernando Mendes. *Evolution of Networks: From Biological Networks to the Internet and WWW*. Oxford University Press, Oxford, 2003.

[23] Mathias Drehmann, Joerg Oechssler, and Andreas Roider. Herding and contrarian behavior in financial markets: An Internet experiment. *American Economic Review*, 95(5):1403–1426, 2005.

[24] Glenn Ellison. Learning, local interaction, and coordination. *Econometrica*, 61(5):1047–1071, 1993.

[25] Eyal Even-Dar and Asaf Shapira. A note on maximizing the spread of influence in social networks. In *Proceedings of the 3rd International Workshop On Internet And Network Economics (WINE)*, 2007.

[26] Jacob Goldenberg, Barak Libai, and Eitan Muller. Using complex systems analysis to advance marketing theory development: Modeling heterogeneity effects on new product growth through stochastic cellular automata. *Academy of Marketing Science Review*, [Online] 1(9), 2001.

[27] Jacob Goldenberg, Barak Libai, and Eitan Muller. Talk of the network: A complex systems look at the underlying process of word-of-mouth. *Marketing Letters*, 12(3):209–21, 2001.

[28] Sanjeev Goyal. Interaction structure and social change. *Journal of Institutional and Theoretical Economics*, 152:472–495, 1996.

[29] Mark S. Granovetter. The strength of weak ties. *The American Journal of Sociology*, 78(6):1360–1380, 1973.

[30] Mark S. Granovetter. Threshold models of collective behavior. *The American Journal of Sociology*, 83 (6):1420–1443, 1978.

[31] Daniel Gruhl, Ramanathan V. Guha, David Liben-Nowell, and Andrew Tomkins. Information diffusion through blogspace. *SIGKDD Explorations (Special Issue on Web-Content Mining)*, 6(2):43–52, 2004.

[32] Nicole Immorlica, Jon Kleinberg, Mohammed Mahdian, and Tom Wexler. The role of compatibility in the diffusion of technologies through social networks. In *Proceedings of the Eighth ACM Conference on Electronic Commerce (EC)*, 2007.

[33] Steve Jurvetson. What exactly is viral marketing? *Red Herring*, pages 110–111, 2000.

[34] Sham Kakade, Michael Kearns, John Langford, and Luis Ortiz. Correlated equilibria in graphical games. In *Proceedings of the Fourth ACM Conference on Electronic Commerce (EC)*, 2003.

[35] Michael Kearns, Michael Littman, and Satinder Singh. Graphical models for game theory. In *Proceedings of the 17th Conference in Uncertainty in Artificial Intelligence (UAI)*, 2001.

[36] David Kempe, Jon Kleinberg, and Éva Tardos. Maximizing the spread of influence in a social network. In *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, 2003.

[37] David Kempe, Jon Kleinberg, and Éva Tardos. Influential nodes in a diffusion model for social networks. In *Proceedings of the 32nd International Colloquium on Automata, Languages and Programming (ICALP)*, 2005.

[38] Masahiro Kimura and Kazumi Saito. Tractable models for information diffusion in social networks. In *Proceedings of Knowledge Discovery in Databases (PKDD)*, 2006.

[39] Ross Kindermann and J. Laurie Snell. *Markov Random Fields and Their Applications*. American Mathematical Society, 1980.

[40] Jon Kleinberg. Cascading behavior in networks: Algorithmic and economic issues. In Noam Nisan, Tim Roughgarden, Éva Tardos, and Vijay V. Vazirani, editors, *Algorithmic Game Theory*. Cambridge University Press, 2007.

[41] David Krackhardt. Structural leverage in marketing. In Dawn Iacobucci, editor, *Networks in Marketing*, pages 50–59. Sage, Thousand Oaks CA, 1996.

[42] Jure Leskovec and Eric Horvitz. Planetary-scale views on a large instant-messaging network. In *17th International World Wide Web Conference (WWW)*, 2008.

[43] Jure Leskovec, Lada Adamic, and Bernardo Huberman. The dynamics of viral marketing. In *Proceedings of the Seventh ACM Conference on Electronic Commerce (EC)*, 2006.

[44] Jure Leskovec, Ajit Singh, and Jon Kleinberg. Patterns of influence in a recommendation network. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD)*, 2006.

[45] Jure Leskovec, Andreas Krause, Carlos Guestrin, Christos Faloutsos, Jeanne VanBriesen, and Natalie Glance. Cost-effective outbreak detection in networks. In *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, 2007.

[46] Jure Leskovec, Mary McGlohon, Christos Faloutsos, Natalie Glance, and Matthew Hurst. Cascading behavior in large blog graphs. In *SIAM International Conference on Data Mining (SDM)*, 2007.

[47] Thomas M. Liggett. *Interacting Particle Systems*. Springer, 1985.

[48] Stephen Morris. Contagion. *The Review of Economic Studies*, 67(1):57–78, 2000.

[49] Elchanan Mossel and Sebastien Roch. On the submodularity of influence in social networks. In *Proceedings of the 39th ACM Symposium on Theory of Computing (STOC)*, 2007.

[50] George L. Nemhauser, Laurence A. Wolsey, and Marshall L. Fisher. An analysis of approximations for maximizing submodular set functions. *Mathematical Programming*, 14:265–294, 1978.

[51] Lionel Pelkowitz. A continuous relaxation labeling algorithm for Markov random fields. *IEEE Transactions on Systems, Man and Cybernetics*, 20(3):709–715, 1990.

[52] Anatol Rapoport and William J. Horvath. A study of a large sociogram. *Behavioral Science*, 6(4): 279–291, 1961.

[53] Matthew Richardson and Pedro Domingos. Mining knowledge-sharing sites for viral marketing. In *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, 2002.

[54] Bryce Ryan and Neal C. Gross. The diffusion of hybrid seed corn in two Iowa communities. *Rural Sociology*, 8:15–24, 1943.

[55] Thomas C. Schelling. *Micromotives and Macrobehavior*. Norton, New York, NY, 1978.

[56] Charles R. Shipan and Craig Volden. Bottom-up federalism: The diffusion of antismoking policies from U.S. cities to states. *American Journal of Political Science*, 50(4):825–843, 2006.

[57] Duncan J. Watts and Steven H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393: 440–442, 1998.

[58] Ivo Welch. Herding among security analysts. *Journal of Financial Economics*, 58:369–396, 2000.